Optimal selection of households for direct marketing by joint modeling of the probability and quantity of response

Otter, Pieter W.; Scheer, Hiek van der; Wansbeek, Tom

CCSO Centre for Economic Research

University of Groningen

CCSO Working Papers October, 2006

CCSO Working paper 2006/06

Optimal selection of households for direct marketing by joint modeling of the probability and quantity of response

Pieter W. Otter
Faculty of Economics, University of Groningen

Hiek van der Scheer
Management Sciences Group, Research International USA, New York

Tom Wansbeek
Faculty of Economics, University of Groningen

CCSO Working Papers
http://www.eco.rug.nl/ccso

Optimal selection of households for direct marketing by joint modeling of the probability and quantity of response

Pieter Otter Hiek van der Scheer Tom Wansbeek*

December 7, 2000

*Otter and Wansbeek: Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands; e-mail: t.j.wansbeek@eco.rug.nl, tel: +31 50 363 3803, fax: +31 50 363 3720. Van der Scheer: Management Sciences Group, Research International USA, 466 Lexington Avenue, New York, USA; e-mail: h.scheer@research-int.com. We would like to thank Penny Spring, Michel Wedel, an Associate editor, and three referees for their helpful comments.

Direct mail selection by joint modeling of the probability and quantity of response

Abstract

We present several methods for the maximization of expected profits when households are selected from a mailing list for a direct mail campaign. The response elicited from the campaign can vary over households, as is the case with fund raising or mail order selling. The decisions taken by the household are (a) whether to respond and, in the case of response, (b) the quantity of response, e.g. the sum donated or the monetary amount of the order. We jointly model both decisions and derive a number of profit maximizing selection methods.

We empirically illustrate the methods using a data set from a charitable foundation. It appears that modeling both aspects of the response yields considerably higher profits relative to selection methods that are based on solely modeling the response probability.

1 Introduction

An important topic in direct marketing is *response modeling*. It involves prediction of some measure of customer response. For direct mail, three kinds of responses can be distinguished, depending on the offer submitted in the mailing. The first kind concerns mailings with fixed revenues (given a positive reply), like subscriber mailings of a magazine, membership mailings, and single-shot mailings offering just one product, for example a book. A second kind concerns mailings where the number of units ordered can vary, e.g. the number of compact discs ordered by direct mail selling or the subscription time (a quarter of a year, half a year, a full year) when a magazine is offered through direct mail. Third, there are mailings with a response that can take on all positive values. This may involve total revenues in purchases from a catalog retailer, or the monetary amount donated to a charitable organization raising funds by mail.

The main purpose of response modeling is to rank the potential *targets*, available on a mailing list, from most to least promising, in order to make a selection. Typically, a test mailing is sent to a relatively small sample of the mailing list. Then a response model is built linking observed response behavior to households characteristics. This model is used to predict the response for the remaining households on the mailing list. By eliminating the least promising targets of the mailing list, the direct marketing company can increase its profits. The mailing list is a crucial component in this process since it has to contain sufficient information about the targets like past purchase behavior and geographic, demographic, and lifestyle variables.

In the recent literature a number of papers have been published that consider target selection; see e.g. Banslaben (1992), Bult (1993), Bult and Wansbeek (1995), DeSarbo and Ramaswamy (1994), Magidson (1988), and Spring et al. (1999). Most of these studies deal with the case of fixed revenues to a positive reply, and hence concentrate on binary choice modeling methods like CHAID, probit, and discriminant analysis; see Bult and Wansbeek (1995) for an overview of target selection techniques. Although this literature recognizes that most direct mail campaigns do not generate simple binary response but rather household specific response, it is hard to find publications that take this aspect into account.

Simon (1987), for example, suggests taking the average amount of purchase from a random sample of the customers on the mailing list over a couple of years, and to use this as the expected value of a potential customer. Then the response to a positive reply is considered fixed as yet, and the response can be modeled again by a binary choice model. Rao and Steckel (1995) suggest using an ordinary least squares (OLS) model to determine the expected revenues, and obtaining the total revenue just as the expected revenues times the probability of response. However, their empirical example is just a binary choice model.

Recently, Bult and Wittink (1996) proposed a method to incorporate the quantity of response. On the basis of the quantity of response in the last year, households are classified in segments a priori. Within each segment the quantity of response is assumed to be this fixed value. Then, for each segment a binary response model is estimated, which is used for selection. Note that (1) this approach is only applicable if information on past purchase behavior is available and (2) it heavily depends on the assumption that this year's quantity of response is equal to that of last year.

Another recent model involving response quantity is presented by Gönül and Shi (1998). Their basic model is a dynamic programming model of the optimal mailing policy of a firm that sells through a catalogue to rational consumers who maximize their utility by deciding whether or not to order. In the of their paper, they show how this model can be adapted to expand the consumers' choice problem to include the monetary value of the purchase. The importance of this model is its combination of two elements, dynamics and monetary value. However, the possible monetary values are taken to be discrete and the state space expands with the number of different monetary values considered. This implies a heavy computational burden and the authors abstain from an empirical application.

The purpose of this paper is to present a unified framework for modeling household specific response in order to optimally select households for a mailing campaign. Our framework specifies the relevant decisions taken by the households. These decisions are (a) whether to respond or not, and, in the case of response, (b) the quantity of response. As is argued by Courtheoux (1987), higher profits can be expected when both decisions are modeled. We present a number of selection methods that take both decisions into account. An empirical illustration shows considerably higher profits when both decisions are modeled explicitly relative to modeling the response probability only.

The paper is structured as follows. In section 2 we present four models that structure both the probability of response and its quantity. For benchmark purposes, we include a model that neglects differential quantity of response. We indicate how these models are used for optimal selection. We give an empirical application in section 3. Section 4 concludes.

2 The models

Consider a direct marketing firm that has to make the decision whether to send a household a mailing or not, given a record of past observations on background variables, e.g. past expenditures, past donations, income, leisure interest, collected in a k-dimensional vector x. In case a mailing is sent to a given household, the profit to the firm, Π, is given by

$$\Pi = AR - c, \tag{1}$$

where R is the random variable given by

$$R = \begin{cases} 1 & \text{if the household responds} \\ 0 & \text{if the household does not respond,} \end{cases}$$

A is the continuous random variable that denotes the monetary value of the purchases, henceforth called the quantity, and c is the cost of a mailing.

The profit function (1) shows the decisions that have to be made by a household. The first decision is whether to respond or not; this is captured by R. The second decision, which is made conditional on the first one, is the monetary value of the purchases. This is indicated by A.

There are various ways to model these decisions. We describe a number of these in increasing order of sophistication, in the following subsections. We also indicate what these models imply for the optimal selection of households.

2.1 The probability model

In this simplest model we neglect differences in the quantity of response and assume that it is $\bar{A}$, say. We denote the inclination to respond by the continuous latent variable R^* that satisfies a linear model,

$$R^* = x'\beta + v, \tag{2}$$

where $v \sim N(0,1)$, independently from x. Whether there is a response or not is indicated by the observed dummy variable R that relates to R^* in the following way: $R = 1$ if $R^* \geq 0$ and $R = 0$ otherwise. Hence the response probability of a household is

$$p \equiv \mathrm{P}(R = 1|x) = \Phi(x'\beta), \tag{3}$$

where $\Phi(\cdot)$ is the standard normal integral. So expected profit satisfies

$$\mathrm{E}\Pi = \bar{A}p - c.$$

It is optimal to select the households with characteristics x such that expected profit, which depends on x through p, is positive. To make this approach operational, we substitute consistent estimators for the parameters. The same remark applies to the other methods to be discussed. In order to keep the notation simple, we do not make a distinction between parameters and their consistent estimates in this section.

The probability model is mainly used as a benchmark here. It reflects a commonly employed practice: select a household if the probability of response exceeds the ratio of cost to average yield.

2.2 The tobit-1 model

From now on we allow for the quantity of response to vary over households. The basic distinction between the models is whether A and R are assumed to be driven by the same structure or by a different structure. That is, AR can be treated as a single variable that depends on x, or A and R can be treated as separate variables. In the latter case, they still depend on x but in a different way. For example, leisure interest could drive the decision to purchase a certain product, while income and social class drive the quantity decision.

If AR is treated as a single variable, then the standard tobit model or tobit-1 model is appropriate. Let $q^* \equiv AR$, then we have

$$q^* = x'\theta + e \tag{4}$$

with

$$q = \begin{cases} q^* & \text{if } q^* > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $e \sim N(0, \sigma_e^2)$. Instead of observing the latent variable q^* we observe the variable q, which is either zero or positive.

The probability of response is now given by

$$p = \Phi\left(\frac{x'\theta}{\sigma_e}\right).$$

Letting $\phi(\cdot)$ be the standard normal density, we obtain

$$\mathrm{E}(q|x, R = 1) = x'\theta + \sigma_e \frac{\phi(x'\theta/\sigma_e)}{\Phi(x'\theta/\sigma_e)}, \tag{5}$$

cf. e.g. Heckman (1979). Hence

$$\begin{aligned} \mathrm{E}\Pi &= \mathrm{E}(q|x, R=1)p - c \\ &= x'\theta\Phi\left(\frac{x'\theta}{\sigma_e}\right) + \sigma_e\phi\left(\frac{x'\theta}{\sigma_e}\right) - c. \end{aligned}$$

Again, optimal selection is based on positive values for this quantity.

2.3 The tobit-2 model

From now on we assume that A and R relate to x by a different structure. We specify the following model. We maintain (2), and assume for the quantity of response:

$$A = x'\gamma + u. \tag{6}$$

For convenience of notation we assume the same x in (2) and (6) but this is innocuous since the elements of β and γ can a priori be set to zero. The assumptions on u and v defines various model specifications.

The tobit-2 model is obtained when we assume

$$\begin{bmatrix} v \\ u \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_u \\ \rho\sigma_u & \sigma_u^2 \end{bmatrix}\right),$$

independent of x. The tobit-2 model has been frequently employed in the econometrics literature for related problems in which there are two decisions to be taken. Applications include job-search (Blundell and Meghir, 1987), insurance claims (Hsiao et al., 1990), charitable donations (Garner and Wagner, 1991), and sales promotions (Chintagunta, 1993, and Krishnamurthi and Raj, 1988).

Under the tobit-2 specification, the probability of response is given again by (3), and (5) is replaced by

$$\mathrm{E}(A|x, R=1) = x'\gamma + \rho\sigma_u\frac{\phi(x'\beta)}{\Phi(x'\beta)}, \tag{7}$$

and

$$\mathrm{E}\Pi = \mathrm{E}(A|x, R=1)p = x'\gamma\Phi(x'\beta) + \rho\sigma_u\phi(x'\beta) - c \tag{8}$$

is the basis for the selection under the tobit-2 formulation.

2.4 The basic TPM

With the tobit-2 model, we assume $\mathrm{E}(u|x) = 0$. An alternative assumption is $\mathrm{E}(u|x, R = 1) = 0$, so now including the conditioning on response. Then

$$a \equiv \mathrm{E}(A|x, R = 1) = x'\gamma, \tag{9}$$

which is simpler than (7). The model is called a *two-part model* (TPM). Applications of the TPM include medical insurance claims (Duan et al., 1983) and vacation expenditures (Melenberg and Van Soest, 1996).

There are several reasons why the TPM may be preferred over the tobit-2 model. First, as was pointed out by Duan et al., 1983, a major advantage of the TPM over the tobit-2 model is that the two parts can be estimated separately. In particular, a consistent estimator of γ is obtained by OLS of A on x. Second, the TPM is far less susceptible of misspecification of the distributional assumptions, as was pointed out by e.g. Hay et al., 1987, and Manning et al., 1987. This Arabmazar and Schmidt (1982) and Goldberger (1981) show that the maximum likelihood estimator of the tobit-1 model suffers from a substantial inconsistency under non-normality and heteroskedasticity of the error term. If this finding carries over to the tobit-2 model, which seems plausible, the results from applying the latter model can only be trusted if the distributional assumptions are correct. Third, the TPM is robust in the following sense. Duan et al. (1983), Hartman (1991), Hay et al. (1987), and Manning et al. (1987) show in extensive Monte Carlo studies that, even if the tobit-2 model is the true model, i.e. if the errors have a bivariate normal distribution, the TPM works very well.

Now, when the TPM is entertained rather than the tobit-2 model, (8) has to be replaced by $\mathrm{E}\Pi = ap - c$. Hence the value a^* of a beyond which a mailing becomes profitable is given by

$$a^* = \frac{c}{p}, \tag{10}$$

which is a hyperbola in the set $\mathcal{R} \equiv \{(a, p)|a \geq 0, 0 \leq p \leq 1\}$ of all possible values of a and p. In order to distinguish this model from the refinement to be discussed now, we call this TPM case the basic TPM.

2.5 The mixture TPM

A refinement of the TPM is obtained by further exploiting the possibility that the distribution of the background variables x differs between respondents and non-respondents. Hence, the distribution of x overall is considered a mixture of two distributions.Assuming γ to be known (again neglecting the difference between parameters and their consistent estimates, and noting that in the TPM γ is simply obtained by OLS), this amounts to stating that the distribution of $a \equiv x'\gamma$ may differ between the two groups. It will then also differ conditional on p. We bring this out by the notation $f_1(a|p)$ for the conditional density of a for respondents and $f_0(a|p)$ for non-respondents (i.e., the quantity if non-respondents would respond), and

$$\lambda(a,p) \equiv \frac{f_0(a|p)}{f_1(a|p)} - 1.$$

When the distribution of a in both groups coincides, $\lambda(a,p) = 0$.

Now, let $\tilde{a}$ some fixed value of a and consider a small set $\mathcal{E}(\tilde{a},p)$ in $\mathcal{R}$, i.e.

$$\mathcal{E}(\tilde{a},p) = \{(a,p)|\tilde{a} - \epsilon/2 \leq a \leq \tilde{a} + \epsilon/2, p\} \cap \mathcal{R},$$

with ϵ small. The incidence of $\mathcal{E}(\tilde{a},p)$ is

$$\begin{aligned} \mathrm{P}(\mathcal{E}(\tilde{a},p)) &= \mathrm{P}(\mathcal{E}(\tilde{a},p)|R=1)\mathrm{P}(R=1) \\ &\quad + \mathrm{P}(\mathcal{E}(\tilde{a},p)|R=0)\mathrm{P}(R=0) \\ &\approx \epsilon\left[f_1(\tilde{a}|p)p + f_0(\tilde{a}|p)(1-p)\right], \end{aligned}$$

with corresponding expected profit

$$\begin{aligned} \mathrm{E}\Pi(\mathcal{E}(\tilde{a},p)) &= (\tilde{a}-c)\mathrm{P}(\mathcal{E}(\tilde{a},p)|R=1)\mathrm{P}(R=1) \\ &\quad - c\mathrm{P}(\mathcal{E}(\tilde{a},p)|R=0)\mathrm{P}(R=0) \\ &\approx \epsilon\left[(\tilde{a}-c)f_1(\tilde{a}|p)p - cf_0(\tilde{a}|p)(1-p)\right]. \end{aligned}$$

For this case, the value a^* of a beyond which a mailing becomes profitable follows from $\mathrm{E}\Pi(\mathcal{E}(a^*,p)) = 0$, which can be solved to yield

$$a^* = \frac{c}{p}\left[1 + \lambda(a^*,p)(1-p)\right]. \tag{11}$$

Notice that, when $\lambda(a,p)$ is assumed bounded, a^* approaches infinity for $p \to 0$ and approaches c for $p \to 1$. This selection rule reduces to the one given by (10) when $\lambda(a,p) = 0$.

We further notice that using the information provided by the conditional density functions $f_0(a|p)$ and $f_1(a|p)$ implicit in $\lambda(a,p)$ will result in a better selection than under the basic TPM. This can be shown as follows. Let

$\bar{a} \equiv c/p$. If $\lambda(a^*, p) > 0$, then $a^* > \bar{a}$, which is clear from (10) and (11). Define

$$\mathcal{A}_1 \equiv \{a | \bar{a} < a < a^*\};$$

all elements in $\mathcal{A}_1$ have negative expected profit and are selected by using the basic TPM but not by using the mixture TPM. If $\lambda(a^*, p) < 0$, then $a^* < \bar{a}$. Define

$$\mathcal{A}_2 \equiv \{a | a^* < a < \bar{a}\};$$

all elements in $\mathcal{A}_2$ have positive expected profit and are not selected by using the basic TPM approach but are selected by using the mixture TPM. Hence, (11) defines a better selection rule than (10), which should result in larger profits. As a suggestive illustration of the difference between the basic TPM and the mixture TPM, we have drawn the 'watershed' curves for the empirical example to be discussed below in figure 1. Both curves are seen to intersect in the middle of the figure. The area between the curves to the southeast of the intersection point represents $\mathcal{A}_1$ and the area between the curves to the northwest of the intersection point represents $\mathcal{A}_2$.

To see that this makes sense intuitively, assume a and p to be positively correlated, and consider a relatively large value of p. Then a^* will be low. Since $f_0(a|p)$ has most 'mass' at lower values of a and $f_1(a|p)$ has most 'mass' at higher values of a, we expect for low values of a^* that $\lambda(a^*, p) > 0$. Hence the curve for the mixture TPM will be above the other curve. The same argument holds vice versa for relatively low values of p.

3 An empirical example

In order to make the methods operational we need estimates of the parameters of the various models. We do so by using the results of a test mailing on a subset of the mailing list. The test mailing identifies respondents and non-respondents, and contains a response quantity for the respondents.

3.1 Implementation

As to implementation, and in particular estimation, we note the following. The probability model is estimated as a standard probit model. For the quantity of response for this model, we use the average quantity of response from the respondents in the test mailing. The tobit-1 model is estimated by maximum likelihood, and the tobit-2 model is estimated with Heckman's two-step estimator (e.g. Amemiya 1985). For the two versions

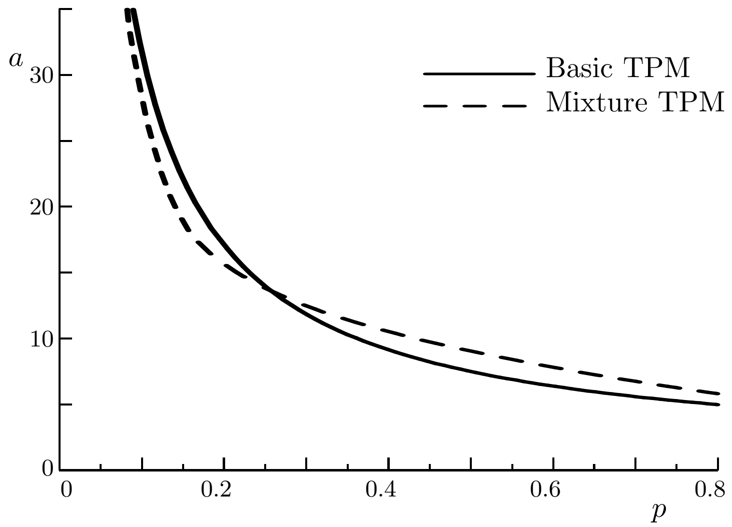

Figure 1: The mailing regions of the approximations.

of the TPM, we estimate γ by OLS on (6) using the data on the respondents, and separately by probit on (2) using the results for respondents and non-respondents.

In order to operationalize the mixture TPM, we use a simple approach and employ the Gaussian kernel, see e.g. Silverman (1986). We make a simplification in the sense that we replace the conditional density functions $f_1(a|p)$ and $f_0(a|p)$ by the marginal density functions $f_0(a)$ and $f_1(a)$ since the estimation of bivariate kernels, needed to estimate the conditional density functions, appeared to be very sensitive to outliers in a and p and to the choice of the smoothing parameter. Then,

$$\hat{f}_k(a) = \frac{1}{n_k h} \sum_{i=1}^{n_k} \phi\left(\frac{x'_{ki}\hat{\gamma} - a}{h}\right),$$

for $k = 0$ and $k = 1$, where $\hat{\gamma}$ is the OLS estimator of A on x; x_{0i} refers to the i-th non-respondent and x_{1i} refers to the i-th respondent in the test mailing. The numbers of non-respondents and respondents are n_0 and n_1, respectively. For the smoothing parameter h we choose $h = 1.06\omega n^{-1/5}$, where ω is the standard deviation of $\hat{a}$ (Silverman 1986, p. 45). Since we estimate two functions we have two smoothing parameters. In order to have only one smoothing parameter, we use the weighted average of these two.

3.2 Data

We illustrate and compare the different methods with an application based on data from a Dutch charitable foundation. This foundation relies heavily on direct mailing. Every year it sends mailings to almost 1.2 million households in the Netherlands.

The data sample consists of 40,000 observations. All households on the list have donated at least once to the foundation since entry on the mailing list. The dependent variable in (6) is the amount of donation in 1991, and in (2) the response/non-response information. The explanatory variables in both models are the amount of money donated in 1990, ditto in 1989, the interaction between these two, the date of entry on the mailing list, family size, own opinion on charitable behavior in general (four categories: donates never, donates sometimes, donates regularly, and donates always). These variables were selected from a database with 58 possible explanatory variables after a preliminary analysis.

The overall response rate $\bar{p}$ is 33.9%, which is rather high but not really surprising since charitable foundations have in general high response rates, and the mailing list only contains households that had donated to the foundation before. The average amount donated $\bar{a}$ was NLG 17.04, and the cost of a mailing c was NLG 3.50.

3.3 Empirical results

In order to obtain a robust insight into the performance of the various methods we use the bootstrap method (e.g. Efron 1982) instead of a single estimation and validation sample. To generate a (bootstrap) estimation sample we draw with replacement 1,000 observations from the data set of 40,000 observations. This sample can be interpreted as the test mailing. We then draw 39,000 observations, again with replacement, to generate a (bootstrap) validation sample. The estimation sample is used to estimate all the parameters (we will not report their estimates since they are not interesting per se for our purpose), and the density functions. Then, for the various methods, we employ the selection rule to compute on the validation sample the actual profits that would have been obtained. We generated 500 bootstrap samples.

Table 1 contains the bottom line results. The last column shows the number of households selected when the various methods are applied. The preceding column gives the profit obtained by this selection by considering

Table 1: Performance of methods

Method	Profit	#Selected
Probability model	89,134	38,165
Tobit-1 model	93,020	34,872
Tobit-2 model	98,764	26,865
Basic TPM	99,028	27,204
Mixture TPM	99,745	25,476

the amounts actually donated by the selected households. Both columns contain the average over the 500 bootstrap replications. If we consider the probability model as the benchmark, then it is obvious that a great gain results from modeling the response quantity.

Surprisingly, the performance of the tobit-1 model is moderate. It definitely performs better than only modeling the response probability but as good as modeling both decisions. This indicates that the assumption of a common structure for the two decisions is too strict. The differences between the other three models is small. It shows, however, that the tobit-2 model does not outperform the basic TPM. This coincides with the fact that the basic TPM performs as least as good as the tobit-2 model when it is used for predictive purposes. The mixture TPM approach does indeed generate more profits than the basic TPM. Given the fact that often a selection has to be made from a mailing list containing millions of households (rather than 39,000) and that the models are used over and over again, small differences may turn out to be big figure in the end.

A further analysis of the performance of the five methods relative to each other is given in table 2. The results may be too suggestive as to a unique ordering of the profits to be obtained by the methods. Since our analysis is based on 500 bootstrap samples we can simply count the number of cases, out of these 500, in which one method yields a higher profit than another method. The table shows that modeling only the response probability and the standard tobit model generally give suboptimal profits. The other approaches more or less equivalent although the OVW approach will on average pay off.

Table 2: Relative performance of methods

	Probability	Tobit model	tobit-2 model	Basic approach
Tobit-1 model	92			
Tobit-2 model	97	93		
Basic TPM	98	95	57	
Mixture TPM	99	95	64	64

Entry (i, j) is the percentage of cases (in 500 bootstrap samples) where method i outperforms method j.

4 Conclusion

We have introduced an approach to joint modeling of response probability and quantity that leads to selection methods that can be applied in practice in a straightforward way. The outcomes of the empirical illustration suggest that adding quantity modeling to probability modeling to current practice can be highly rewarding. Even the simplest approach to joint modeling can add significantly to profitability.

There are various limitations to the paper that should be addressed in future work. The results of the empirical illustration are highly evocative. The findings suggest that modeling only response probabilities, the focus of much work in target selection, misses a dominant feature in striving for optimality; the gain to be had when quantities are taken into account is large. This may be a specific result, and we do not claim generality. The example concerns charitable donations, and the picture might be qualitatively different when the proposed methods are applied to the other leading case where response is household specific, money amounts involved in mail order buying.

Another limitation is the continuous nature of the quantity variable. As mentioned in the opening paragraph of the Introduction, the response may be of a discrete nature, calling for an ordered probit or Poisson regression model. We did not elaborate this more complicated model but guess that the extension is relatively straightforward and will lead to a better selection.

As a final issue, our approach is limited in the sense that the underlying model is static and does not take behavior over time into account. This issue has two aspects. In the first place, the behavioral model should be improved into a panel data model where a central role is played by the individual effect; household response vis-à-vis direct mailing will have a strong, persistent component largely driven by unobservable variables. The other aspect concerns the optimality rule to be applied by the direct mailing organization, which is essentially more complicated than in the one-shot, static case considered by us. The search is for a model that combines dynamics and monetary value (in discrete amounts), as considered by Gönül and Shi (1998), with the simple and continuous treatment of amounts as presented in the present paper.

References

Amemiya, Y. (1985), "Instrumental variable estimator for the nonlinear errors-in- variables model", *Journal of Econometrics*, **28**, 273–289.

Arabmazar, A. and P. Schmidt (1982), "An investigation of the robustness of the tobit estimator to non-normality", *Econometrica*, **50**, 1055–1063.

Banslaben, J. (1992), "Predictive modeling", in E.L. Nash, editor, *The Direct Marketing Handbook*, McGraw-Hill, New York, 620–636.

Blundell, R. and C. Meghir (1987), "Bivariate alternatives to the tobit model", *Journal of Econometrics*, **34**, 179–200.

Bult, J.R. (1993), "Semiparametric versus parametric classification models: An application to direct marketing", *Journal of Marketing Research*, **30**, 380–390.

Bult, J.R. and T.J. Wansbeek (1995), "Optimal selection for direct mail", *Marketing Science*, forthcoming.

Bult, J.R. and D.R. Wittink (1996), "Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising", *International Journal of Research in Marketing*, **13**, 215–226.

Chintagunta, P.K. (1993), "Investigating purchase incidence, brand choice and purchase quantity decisions of households", *Marketing Science*, **12**, 184–208.

Courtheoux, R.J. (1987), "Database modeling: Maximizing the benefits", *Direct Marketing*, **3**, 44–51.

DeSarbo, W.S. and V. Ramaswamy (1994), "CRISP: Customer response based iterative segmentation procedures for response modeling in direct marketing", *Journal of Direct Marketing*, **8 (3)**, 7–20.

Duan, N., W.G. Manning, C.N. Morris, and J.P. Newhouse (1983), "A comparison of alternative models for the demand for the medical care", *Journal of Business & Economic Statistics*, **1**, 115–126.

Efron, B. (1982), *The jackknife, the bootstrap and other resampling plans*, Society for Industrial and Applied Mathematics, Philadelphia.

Garner, T.I. and J. Wagner (1991), "Economic dimensions of household gift giving", *Journal of Consumer Research*, **18**, 368–379.

Goldberger, A.S. (1981), "Linear regression after selection", *Journal of Econometrics*, **15**, 357–366.

Gönül, F. and M.Z. Shi (1998), "Optimal mailing of catalogs: a new methodology using estimable structural dynamic programming models", *Management Science*, **44**, 1249–1262.

Hartman, R.S. (1991), "A Monte Carlo analysis of alternative estimators in models involving selectivity", *Journal of Business & Economic Statistics*, **9**, 41–49.

Hay, J.W., R. Leu, and P. Rohrer (1987), "Ordinary least squares and sample-selection models of health-care demand", *Journal of Business & Economic Statistics*, **5**, 499–506.

Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, **47**, 153–161.

Hsiao, C., C. Kim, and G. Taylor (1990), "A statistical perspective on insurance rate making", *Journal of Econometrics*, **44**, 5–24.

Krishnakumar, J. (1988), *Estimation of simultaneous equation models with error components structure*, Springer Verlag, Berlin.

Magidson, J. (1988), "Improved statistical techniques for response modeling", *Journal of Direct Marketing*, **2 (4)**, 6–18.

Manning, W.G., N. Duan, and W.H. Rogers (1987), "Monte Carlo evidence on the choice between sample selection and two-part models", *Journal of Econometrics*, **35**, 59–82.

Rao, V.R. and J.H. Steckel (1995), "Selecting, evaluating, and updating prospects in direct mail marketing", *Journal of Direct Marketing*, **9 (2)**, 20–31.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Simon, J.L. (1987), *How to Start and Operate a Mail-Order Business*, McGraw-Hill, New York.

Spring, P.N., P.S.H. Leeflang and T.J. Wansbeek (1999), "The combination strategy to optimal target selection and offer segmentation in direct mail", *Journal of Market Focused Management*, **4**, 187–203.